半双工全光千兆位以太网网卡的软硬件设计

摘要

随着信息技术的飞速发展，如今的用户有了非常强大的桌面计算能力；网络多媒体和网络计算尤其是科学与工程计算的应用不断增加；对网络带宽的需求也不断增长，有力的推动了网络技术的不断发展。而以太网（EtherNet）就是其中的佼佼者。

自 1973 年，Xerox 公司组建了第一个实验性的以太网以来，它已经经历了近三十年市场的考验，得到了广大用户和技术人员的青睐。不断推出新的标准；传输介质不断更新；数据传输速率也不断提高。到 1998 年，IEEE 正式提出了编号为 802.3z 的千兆位以太网标准 1000BASE-X，采用光纤作为传输介质，使用高达 1250Mbps 的基带传输速率，为用户提供了宽阔的数字高速公路。

作者坚信，随着网络技术的发展，光纤一定会代替铜缆成为网络的主流传输介质，目前光纤本身的价格也可以为客户所接受；但是影响光纤介质的局域网推广普及的一个重要因素，就是目前市面上的光网络都采用了全双工交换机，而光纤接口的交换机尤其是千兆位的交换机价格很高，多用于校园网或城域网；只有极少数大型企业能够负担，影响了光纤网络的部署。

为了降低网络成本，本文提出了一种工作组级的基于无源星形耦合器（PSC）的半双工全光千兆位以太网，即以星形光纤耦合器取代光纤接口的千兆位交换机，而以网卡实现半双工的工作方式，对以太网来说，也就是实现载波监听、多路访问/冲突检测（CSMA/CD）协议。这本来是以太网的基本协议，它的基本原理在于所有站点争用共享的网络带宽，任一时刻网络传输介质上只能有一个合法的载波传输，否则即为发生冲突，所有站点都应后退等待一段时间，再重新争用介质。当网络容量增大，传输速率提高时，这种方式会严重影响到网络性能。但是，在目前情况下，出于以下几个方面的考虑，使用 CSMA/CD 方式的半双工千兆位网络也有它的发展空间：

第一、在目前的情况下，有大量的小型局域网存在；但是又要求高的传输速率，如工作室、实验室等工作组级的网络，尤其是现在十分常见的多媒体应用、图形图像处理、虚拟现实与广告渲染等方面。它们的并发传输不多，以突发传输为主，但传输的数据量大，网络的速度会明显影响到工作效率，有向千兆位升级的需求。

第二、采用该方案的网络很容易向全双工交换式的网络升级，该网卡采用美国国家半导体（National Semiconductor）公司的 DP83820 作为网络控制芯片，可以同时支持半双工与全双工操作，只需改动软件。今后一旦交换机的价格下降到可接受的水平，直接用光纤交换机替代该网络的星形耦合器，并升级驱动程序，即可无缝的过渡到全双工交换式网络。

第三、成本低，利于光纤网络部署的推广。目前市面上的 4×4 的光纤交换机价格大约在人民币 28,000 元左右，8 口的更高达 50,000 元，而一个 4×4 的星形耦合器不过人民币 3000 元左右，8×8 的也不过 5000 元。其性能也远超过价格相当的 100M 采用 5 类 UTP 的交换机。而且我国的网络通信发展十分迅速，直接部署光纤网络，方便于用户的升级，更有利于保护用户的投资。

以太网对应了 OSI 开放系统模型（7 层网络模型）的最下两层：数据链路层和物理层，并将其细分为一些功能子层，包括数据链路层的 LLC（逻辑链路控制）子层和 MAC（介质访问控制）子层，物理层的 PCS（物理编码子层）、PMA（物理介质附加子层）和 PMD

（物理介质相关子层）。其中，逻辑链路控制（LLC）由操作系统和网卡驱动程序来实现，介质访问控制（MAC）和它下面的物理层都由网卡实现。

目前市面上没有半双工的千兆位光纤网卡出售，本文主要的工作是自行设计半双工的千兆位光纤网卡的硬件与驱动程序。该网卡主要由 3 块集成的芯片完成其功能，分别是 i）网络控制主芯片，对应于以太网的 MAC 子层，主要完成 CSMA/CD 介质访问协议，管理片上集成的发送和接收缓冲区，并提供和主板 PCI 总线的接口；ii）SERDES（串行解串行化器）芯片，对应于以太网的 PCS 和 PMA 子层，主要完成 8B/10B 编码并将 10 位并行的数据转换为串行数据，在接收端完成相反的功能；iii）光纤收发器，完成串行数据的光电转换功能。由于有大量的高速数字信号，在设计过程中，信号完整性是最主要的问题，所有板上的布线都遵循了一般的高频数字电路的设计原则，在拐角、长度限制、并行距离、阻抗匹配等方面都做了考虑；1.25Gbps 的差分串行线使用了微带传输线结构；高频电路都用 Protel 99 进行了信号完整性分析。此外，电源滤波电路也都按供应商推荐的方法进行了仔细的设计。目标是同时支持半双工和全双工的操作，只需升级驱动程序。

驱动程序运行于内核环境下，拥有对系统的完全的控制，所以其设计不同于一般的 Windows 应用程序，不能使用 C 和 C++提供的标准程序库，而只能使用微软提供的标准内核例程。内存的分配、使用和回收，自旋锁的获取和释放，内核例程所运行的特定的硬件中断级等，都需要特别的注意。从而既保证驱动程序的稳定运行，又保证占用最少的系统资源。微软对网络驱动程序又有其专门定义的程序结构，用网络驱动程序接口规范库（Ndis）来定义了全部的接口，并提供系统调用的封装。网络驱动程序的主要内容，也就是具体完成在驱动程序初始化时建立的_NDIS_MINIPORT_CHARACTERISTICS 结构所规定的所有例程。

论文共由四章构成：

第一章介绍了以太网的基础知识、其发展历程与现状；简要介绍了 OSI 开放系统模型与千兆位以太网标准参考模型，以及 IEEE802.3z 标准；最后提出了一种半双工全光千兆位以太网的结构。

第二章介绍了网卡的硬件设计。主要包括网卡的工作原理，介绍了设备与元器件的选型；给出高频数字电路的设计原则；数字电路的电源与地平面的设计方法；计算并设计微带传输线；最后给出电路图与 PCB 板布线图。

第三章描述了网卡驱动程序（For Windows2000）的设计，介绍了 Windows 2000 驱动程序的基本结构；网络驱动程序接口规范库（Ndis）支持的微端口（Mimiport）驱动程序的构造，并给出标准 C 语言的程序源代码。

第四章展望了该网卡的应用前景，并对它的升级方向作了一些探讨。希望能够将该网卡投入实用，并为本系的光通信实验室的建立积累一些材料，包括第一手的软硬件设计的经验，常用的设备与元器件的资料等。

关键词：半双工千兆位以太网， 以太网网卡， 驱动程序

Abstract

With the development of information technology, we have had powerful desktop computer. Applications of network multimedia and network computing especially in engineering and science become more and more. The need of network bandwidth is more urgent. These make the network technology get fast development. The Ethernet is a star of those networks.

Since the Xerox make up the first experimental Ethernet in 1973, it had gone through the baptism of market and gotten the trust of many users. New standards are presented continuously, link media type is developing and the line speed is updating continuously. In 1998, the IEEE802.3z Gigabit Ethernet standard was ratified, it use fiber as it's link media, line speed was upgraded up to 1,250 Mbps. It offers the users wide digital highway.

The author believe that copper cable will be replaced by fiber as the development of network, but now most fiber LAN use a switch with fiber interface to get full-duplex operation, and the most important obstacle in deploying fiber LAN is the high price of the switch. Only some big company can afford the fiber switch and the switches often work in MAN.

In order to lower the costs of Gigabit optical LAN, this paper presents a workgroup optical network base on passive star coupler (PSC). It use a PSC to replace the switch, and implement the CSMA/CD in the network interface card (NIC). When the throughput grows and transfer speed is raised, CSMA/CD will lower the network's performance, but think about such truths below, the half –duplex Gigabit optic Ethernet will have some market space.

First, there still exist large amount of small-size local network nowadays; but they require high transmit speed, such as studio, lab etc, especially multimedia, graph and image processing, virtual reality and advertising. Their transmissions are rarely simultaneous, but mainly paroxysmal.

Second, network that based on this solution is convenient to be upgraded to full duplex. This NIC uses DP83820, which is developed by National Semiconductor of American, as the network control chip and could support both half duplex and full duplex just only to modify the software. Once switches prices descend to a acceptable level in the future, this network can be easily transited to full duplex network through replacing the star coupler with fiber switch, only need upgrading driver.

Third, its low cost can push forward the spread of fiber networks. Current market value of 4×4 fiber switch is about 28,000 Yuan, and the 8×8 50,000 Yuan, but a 4×4 start coupling costs only 3000 Yuan, and 8×8 5000 Yuan. Its performance is far better than the same-price 100M switches which adopt Category 5 UTP. Network communications are developing rapidly in our country, deploying fiber network directly is convenient for users to upgrade their systems and even more benefit to protect users investments.

Ethernet maps the two lowest layers of OSI model(Seven-layer Module):Data Link layer and physical layer and divides them to some function sublayers including PCS(Physical Coding Sublayer), PMA(Physical Medium Attachment) and PMD(Physical Medium Depend) in physical layer. In these sublayers, Logical Link Control(LLC) is implemented through operation system , NIC driver software and Medium Access Control(MAC) and the physical layer below it are implemented by the NIC.

No half duplex Gigabit fiber NIC is sold in the market now. The main task of this paper is to design the hardware of half duplex Gigabit fiber NIC and develop its driver independently. This card largely depends on three integrate chips to fulfill its function: 1) NIC control main chip,

corresponding the MAC sublayer of Ethernet, to realize CSMA/CD media access protocol, manage the sending and receiving buffers integrated on the chip and provide motherboard PCI interface. 2)SERDES(SERializing and DESerializing)chip, corresponding PCS and PMA sublayers in Ethernet, mainly to complete 8B/10B coding and convert 10 bits parallel data to serial data, and convert them again at the receiving end. 3) fibre transceiver unit, completing light-electrical conversion of serial data. Because of large amount of high speed digital signals, integrality of these signals becomes the most important problems in the design. All the lines on the board follow the general principle of high frequency data circuit, factors such as corner, length limit, parallel length and impedance matching etc, are taken into account; 1.25Gbps differential serial lines adopt the micro-band transmit line architecture; The signal integrity of high speed circuits are checked by Protel 99. Besides, power-filter circuits is carefully designed following the way recommended by the providers. Its goal is to support both full duplex and half duplex, and only need to upgrade the software when upgrading the NIC from half duplex to full duplex.

Device driver is running in the kernel environment and own the right to control the system completely, so, its design is different to general Windows application programs. Instead of utilizing the standard program library provided by C and C++, it can only use the standard kernel functions provided by Microsoft. Problems such as the allocation, utilization, callback of memory, obtaining and release of spinlock, special hardware interrupt level needed by kernel process, should be pay more attention to make sure that the driver can run stably and occupy the least system resource at the same time. Microsoft specially defines the programs architecture of network driver programs: it defines all the interface through network driver interface standard(Ndis) and provides the encapsulation used by system. The primary content of network driver are all the functions prescribed by the _NDIS_MINIPORT_CHARCTERISTICS architecture when initializing the driver.

This thesis consisted of four chapters:

Chapter One: Introduce the basic knowledge, history and status in quo of Ethernet; simply describe the OSI model and KM bits Ethernet standard reference model and IEEE802.3z standard; bring up the architecture of half duplex optical Gigabit Ethernet finally.

Chapter Two: Introduce the hardware design of the NIC, mainly including the work principle of the NIC, the selection of equipment and component; design principle of high frequency circuit; power and ground design method of data circuit; compute and design micro-band transmission; and supply circuit diagram and PCB board lines deploying diagram finally.

Chapter Three: describe the design of NIC driver(For Windows2000), introduce the basic structure of Window2000 driver programs; architecture of Mimiport driver programs supported by network driver interface standard, and provide the source code of the program in standard C language.

Chapter Four: foresee the future of this NIC and discuss its upgrading direction. We hope that NIC can be used practically, and to accumulate some materials for the Fiber Communication Lab of the Department of Mechanical and Electrical Engineering, which including the first hand hardware and software design experience, data of equipment and component used frequently.

Keyword: half-duplex Gigabit Ethernet, Ethernet Network Interface Card, NIC Driver

目录：

第一章 半双工全光千兆位以太网简介

1.1 以太网基础知识

1.1.1 CSMA/CD 简介

以太网起源于一个实验网，是 Xerox 用来支持其 Palo Alto 研究中心的“Alto”个人计算机之间的 3Mbps 互联。以太网 CSMA/CD（载波监听，多路访问/冲突检测）协议有 3 条明确的规定：

• 载波监听：所有运行的站必须连续监听网络。如果出现一个信号载波，表明网络出于忙状态；如果不出现载波，则表示网络安静，可认为不忙。

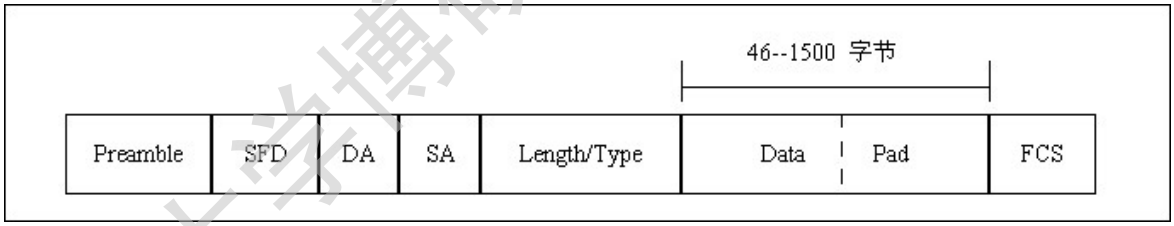
• 多路访问：任何站只要检测到网络已空闲一小段时间，都可以发出一帧数据。

• 冲突检测：如果两个以上的站点在同一段时间内发送帧，则发送的位流会相互干扰（冲突），则这些帧都被破坏，每个站都必须能够检测发送期间的位流冲突。

在线缆上可以传输基带或宽带信号，现代以太网都使用基带传输。早年对宽带选项做了规定，但是没有得到市场的认可。

1.1.2 以太网帧结构

帧（Frame）是以太网物理介质上传输的数据段的单元，任何一个站的数据发送与接收，都是以帧为单位。基本的以太网帧结构如下：



SFD＝帧起始定界符 DA＝目标地址 SA＝源地址 FCS＝帧校验序列

图 1.1 基本 MAC 帧格式

该格式中每个字节的顺序是低位在左，高位在右，只有 FCS 段例外，FCS 作为一个 32 位字段（不是 4 字节）是高位在左，低位在右。

• 预同步码（Preamble）：预同步码的目标是允许物理层在接收到实际的帧起始定界符之前检测载波，并且与传入的帧达到稳定同步。其构成为 7 字节的 10 交替序列，最后一个是 0。

• 帧起始定界符（Start-Of-Frame Delimiter ,SFD）:用以识别帧的起始位置，并同步格式字节边界，SFD=10101011。

• 地址字段（Address）：地址字段包括目标地址和源地址，都是完全的 6 字节地址。第一位（最左边）为 0 指示该地址是一个单个地址，为 1 表示后面是一个组地址。第二位 0 表示全局管理地址，1 为本地管理地址。后跟统一分配给各个生产厂商的 46 位地址。每一

块网卡都有它自己的地址。全 1 的地址为广播地址，代表网络中的所有站。

• 长度/类型字段：指出数据字段中的数据字节数和帧的类型。如果长度/类型字段值<1500,则数据字段包含有数据且其字节数等于长度/类型字段的值；如果长度/类型字段的值大于 1536（0x0600），那么该帧是一个类型帧，其值表示该帧的类型。

• 数据和填充字段（Data And Pad）：数据字段的所有数据对传输的帧是透明的。数据字段的长度可以是不大于 1500 的任意值。但是，如果长度小于 46，必须增加一个填充（Pad）将其扩展，以使其达到图 1.1 与图 1.2 所示的最小值。但是这个填充不会改变长度/类型字段中的长度值。

• 帧校验序列（Frame Check Sequence, FCS）：是对从目标地址开始的，DA+SA+Length/Type+Data+Pad 计算生成的 32 位校验序列。在接收时，将对以上各段重新生成 FCS，并与接收到的 FCS 比较，从而可以知道传输错误。

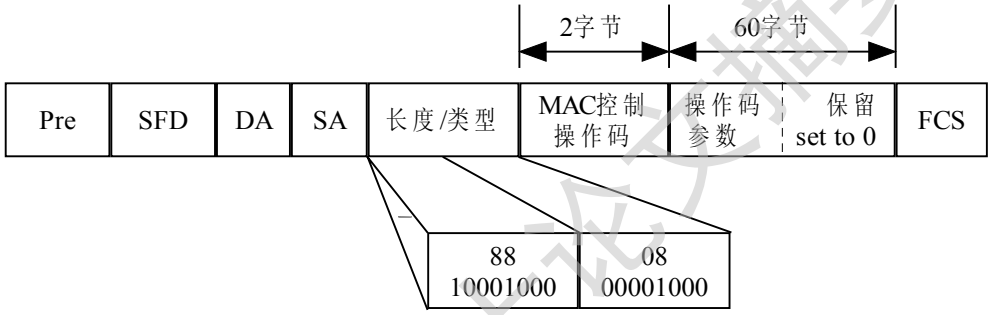


图 1.2 MAC 控制帧格式

控制帧的格式如图 1.2。控制帧目前只定义了 PAUSE（暂停），用于提请链路的另一方暂停帧传输。当发送 PAUSE 帧时，必须使用全局的 16 进制目标地址 01－80－C2－00－00－01。长度/类型字段将用于标识尚未定义的控制帧。

• 控制参数和保留字段：控制参数可能包含 0 个，1 个或更多的参数，此参数为操作码专用。如果控制参数小于 60 字节，就加上保留字节（设为 0）来扩展参数字段，从而使（参数＋保留）的长度等于 60 字节。PAUSE 帧使用一个指示暂停时间的参数，以 512 位时间（bit time）为单位，延时的值可以是 0 到 65535 个 512 位时间。

VLAN 标记的帧格式如图 1.3 所示：

802.1Q 所规定的 VLAN（虚拟网）标记类型被赋予唯一的值 0x81－00。而标记控制信息字段包括以下三种指示符：用户优先级指示了传输的优先级别（0～7，7 为最高优先）；CFI 用来指示数据字段中有没有路由选择信息字段（RIF）,VLAN 标识符是用于帧传输的 VLAN 标识号。

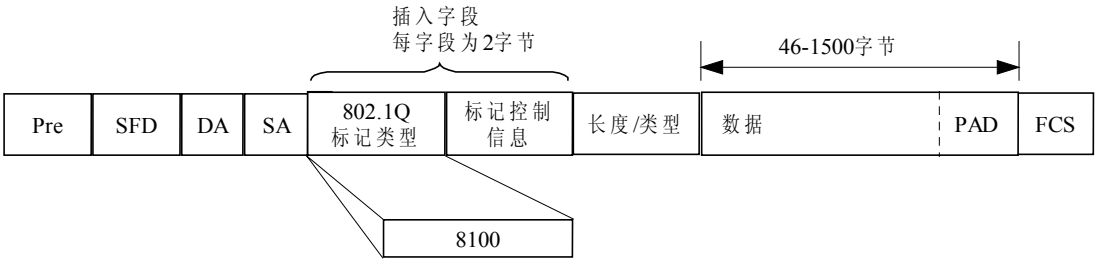


图 1.3 VLAN 标记的 MAC 帧格式

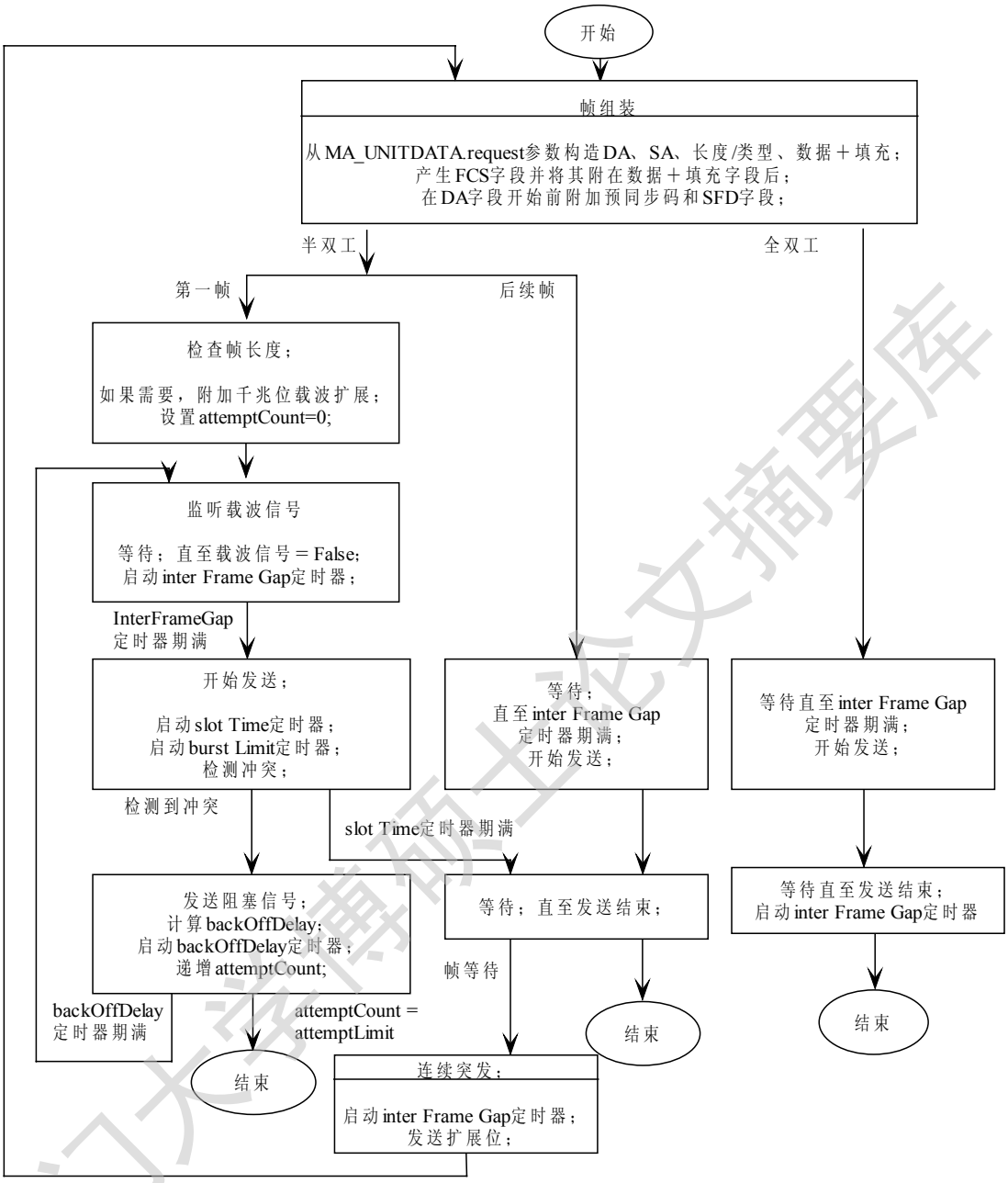


图 1.4 千兆位以太网帧发送流程

1.1.3 帧发送

只要在一定时间内监听不到**载波**信号，CSMA/CD 协议允许任意站开始发送一帧。如果在同一空闲期间 k 和 q 站都开始发送，这两个位流必然相互**冲突**，当发送站收到的位流与正在发送的位流不同时，该站就知道发生了冲突，它就发送一个**阻塞信号**，使得其他站也知道发生了冲突，然后停止发送。要指示出成功的帧发送，要求一个帧长度要长到足以保证在整帧送出之前能够检测到冲突。最坏的情况发生在当两个站在网络中最长路径的两端的时候，直到第一个站发送的帧到来前第二个站还未开始发送。因此从发送开始到第一个站检测到冲突这段时间是整个网络最大传输延迟的两倍，这个周期被称为**冲突窗口**。所有这些可能的站对之间的路径集合称为**冲突域**。冲突域中最长路径的长度为**冲突域直径**。当冲突窗口期满时，发送站可以认为该帧会被正确接收。在冲突域中的其他所有站都将会监听到载波，并一直等到网络再次回到空闲状态以后再尝试发送数据。为了降低发送站间出现冲突的概率，每个冲

突站将其重发延迟一个随机选择的时间。如果额外冲突仍然发生，则将每个相继冲突之后选择的随机等待时间延长，直到重传成功或该站达到为连续发送尝试次数设立的极限。

以太网帧发送的功能性流程图见图 1.4。

1.1.4 帧接收

帧的接收过程如图 1.5。

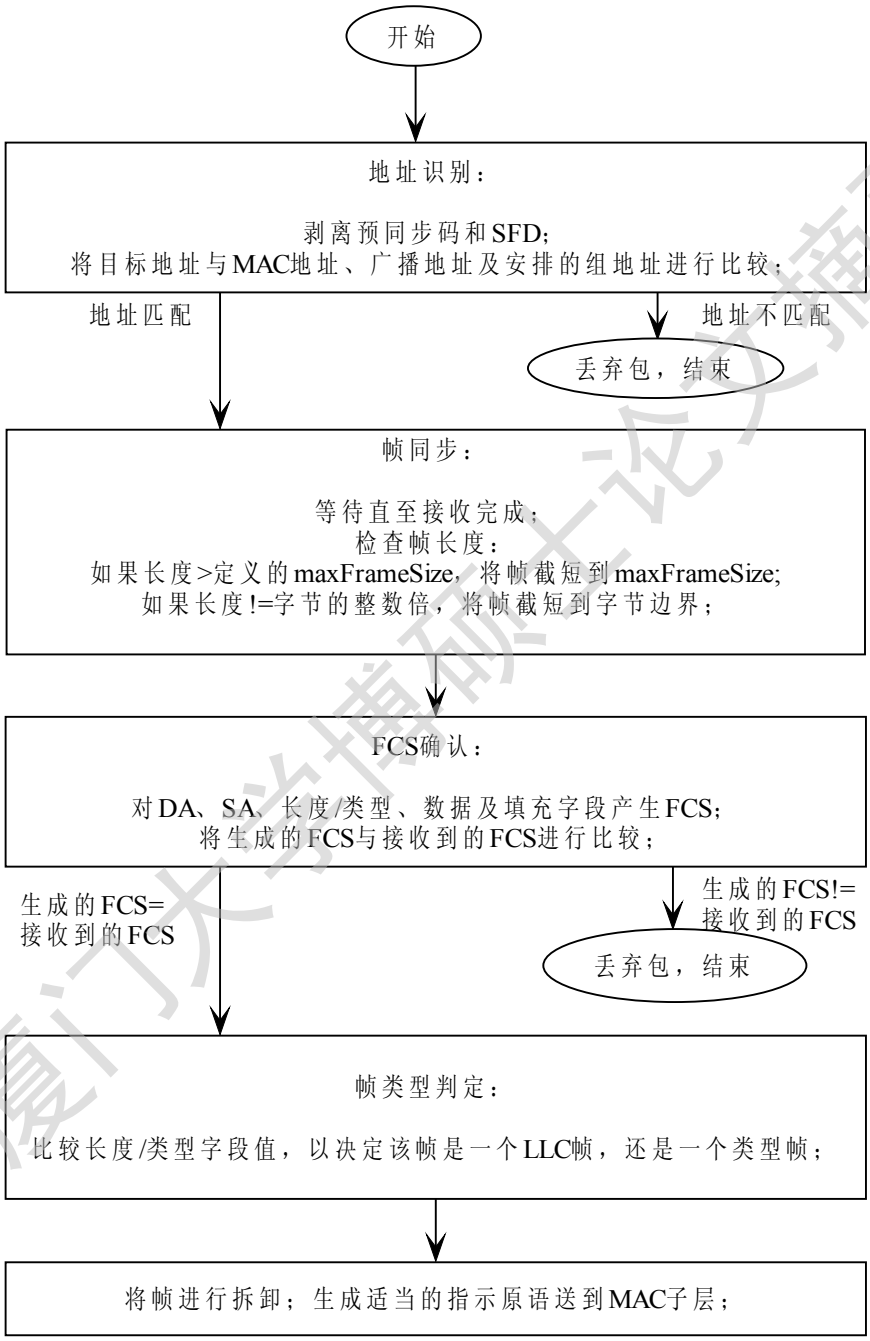


图 1.5 千兆位以太网接收流程

1.1.5 以太网的基本组成

以太网本身由网络站（在标准中称为 DTE，即数据终端设备）、中继器、网桥及其网络互联介质（总线和链路）组成。

网络站（DTE）：DTE（Network Station）通常是个人计算机、工作站或像打印机及存储设备这样的网络服务器。每个 DTE 配备所以必需的硬件和软件，以便实现 CSMA/CD 协议的功能要求。一个网络中最多可以有 1024 个 DTE。

中继器（Repeater）：中继器允许网络的长度和拓扑结构扩展超出单条链路段的极限。中继器是局域网（LAN）链路段间的有源连接部件，但从网络控制的观点来看它本质上还是无源的。中继器负责以太网 CSMA/CD 协议规则中的两项操作：载波监听和冲突检测。中继器检测来自链路段的传入帧（载波监听），恢复接收到的位流的信号特性，然后将该位流中继（传送）到其他连接的链路段。当检测到任一链路段上的冲突时，中继器就将一个阻塞信号发到相关的两个链路段上。中继器不是 DTE，不计入站的总数。

网桥（Bridge）：网桥是一种特殊的 DTE，它们常用来将一个 LAN 与另一个 LAN 互联。根据需要，它们为产生于一个网络并正向另一个网络中发送的帧提供协议和速度变换。连至网桥的以太网就如同其他以太网 DTE 一样。

网桥检查所有传入帧的传输差错，它只传递没有错误的帧。如果另一个相连的 LAN 也是一个以太网网络，那么网桥以另一个 LAN 的传输速度缓存并重发传入帧。如果另一个 LAN 不是以太网网络，则网桥负责把传入帧变换成适合那个 LAN 协议的格式。网桥学习并存储连接的两个网络上的每个站的源地址，只转发有正确的已知地址的帧。

总线和链路（Bus and Link）：在一个以太网网络中连接 DTE、中继器和网桥的物理介质称为总线或链路，主要依据网络中的介质类型、传输速度和线缆位置决定。目前的介质类型包括同轴电缆、双绞铜缆和光缆。以太网网络中任两个站间可以有一条，也只能有一条传输路径。

1.2 以太网的发展与现状：

1973 年，Xerox 公司的 Palo Alto 研究中心为了互联其“Alto”个人计算机，组建了一个实验性的网络。用一根共享的电缆作为物理介质将 3Mbps 的数据流送向所有的节点。他们将这根共享的电缆称为“以太”（Ether）。以该网络为基础，1980 年，由 DEC，Intel 和 Xerox 公司共同提出了关于以太网协议标准的正式建议，建议在一个共享的介质上实现 10Mbps 基带传输的网络。这个网络可互联最多 1024 个节点，最大的节点间距离为 2.5 公里。

1985 年，以太网协议首次成为正式的 IEEE 标准，此后又有大量的修订版和增补版，它们都在 IEEE802．3 或 ISO／IEC8802-3 的号码系列之下，并且根据可选电缆类型和传输方法，规定了版本名，一开始是：（传输速率）（传输方式）（网络直径），后来改为（传输速率）（传输方式）－（介质类型）。如：

10BASE5 －－10Mbps 基带传输，５００米同轴电缆；

10BASE-T －－10Mbps 基带传输，双绞线；

100BASE-FX－－100Mbps 基带传输，光纤；

1000BASE-LX－－1000Mbps 基带传输，长波光纤。

1985 年，10BASE5 规范作为 IEEE-802.3---1985 发布。以最长 500 米的粗同轴电缆为传输介质，因为粗缆的直径太大而且很硬，这种 LAN 相当昂贵并难以安装。但是，它证明了以太网协议是可行的，而且网络配置比较方便。

1990 年，10BASE-T 的推出是以太网成功的里程碑。它的成功归于多端口中继器和结构化布线的使用。多端口中继器（又称集线器 HUB），作为星型布线的中心环节，加上结构化电话布线，使得 10BASE-T 网络成本低廉,可靠性大大提高,网络配置十分灵活,而且其

网络管理与故障排除都很简单。立刻得到了市场的认可， 1995 年，全世界销售的局域网端口中 10BASE-T EtherNet 占 87%，从而也为以太网后来的持续发展打下了良好的基础。

1993 年底开始发展的快速以太网在 1995 年推出了 100BASE-X 标准。它与 10BASE-T 有着相同的基础：以太网帧格式，多端口中继器，网桥，和结构化布线；并将传输速率提高了 10 倍。因此也有一些改动，首先是电缆，100BASE-T4 使用 4 对 3 类或 5 类 UTP（非屏蔽双绞线），100BASE-T 使用 2 对 5 类 UTP 或 STP（屏蔽双绞线）。其次由于传输速度的提高，冲突域直径（或网络直径）必须相应减小，使用单个中继器的 100BASE-T 其网络直径不超过 200 米。

1994 年，针对快速以太网的上述弱点，交换式快速以太网出现了。在不改变传输速度、DTE 接口类型，并不降低冲突域直径的情况下，可以得到 2 倍于快速以太网的有效带宽。此后还出现了 10M/100M 通用端口交换机，为高通信量端口提供 100Mbps 的速率，为普通多数用户提供 10Mbps 的端口；并允许光纤链路增大到 400 米。

1998 年，IEEE 推出了 802．3z 千兆位以太网标准 1000BASE-X。该标准使用了与以前的以太网相同的帧格式，规定了长波光纤、短波光纤和短铜跳线三种介质上的 1000Mbps 的传输。1999 年又推出了名为 IEEE802．3ab 的标准 1000BASE-T，提供了在 100 米的 4 对 5 类 UTP 上传输 1000Mbps 数据的方法。为广大以太网用户提供了无缝的升级手段，进而成为宽带的 LAN、校园网主干乃至城域网的解决方案。

1.3 千兆位以太网与 OSI 标准模型

1.3.1 ISO/OSI 网络参考模型与以太网分层模型：

ISO/OSI 的七层模型是 1977 年提出的，以后的所有计算机网络都在这个框架下构建。这七层分别是：

应用层：实现用户的网络接口功能。通常处理如登录和口令鉴别，提供目录服务，在命名的等层间建立联系，执行报文处理，并处理并发事件和错误恢复。

表示层：协商传输语法，执行任何需要的数据转换，如字符集和文本流的翻译和任何允许两个节点相互交互所需的加密/解密。

会话层：建立和管理通信对话。此层将地址映射为名字，决定一个通信何时开始和结束，以及控制每个设备何时传输和传输多长时间。

传输层：该层是与穿过网络的数据传输直接相关的最高层。它负责保证传输和接收设备之间的可靠通信。此层处理端到端差错检测和恢复，提供流量控制，监视服务质量，拆卸/重装会话数据和记录失序分组等。

网络层：建立网络路由选择，翻译地址，识别分组的优先级，以适当的优先级次序发送分组，并处理网络互联。当数据太长，超过局域网协议允许的最大帧长度时，网络层也可以拆卸及重装传输层的数据。对以太网很重要的一种网络层协议是网际协议（IP）。

数据链路层：提供对 LAN 的访问，保证数据通过某条链路的可靠传输。为了完成传输，此层将数据帧格式化，为差错检测生成帧校验序列，并启动传输。在帧接收时，此层过滤目标地址，检查传输差错和帧格式差错。数据链路层提供传输差错提示，但不能从错误中恢复。错误恢复是高层协议（如传输层）的职责。

物理层：负责对连到通信介质的所有电气，光学和机械要求。此层提供位流编码/译码，同步，时钟恢复和发送/接收。

IEEE802.3 定义的以太网参考模型如图 1.6

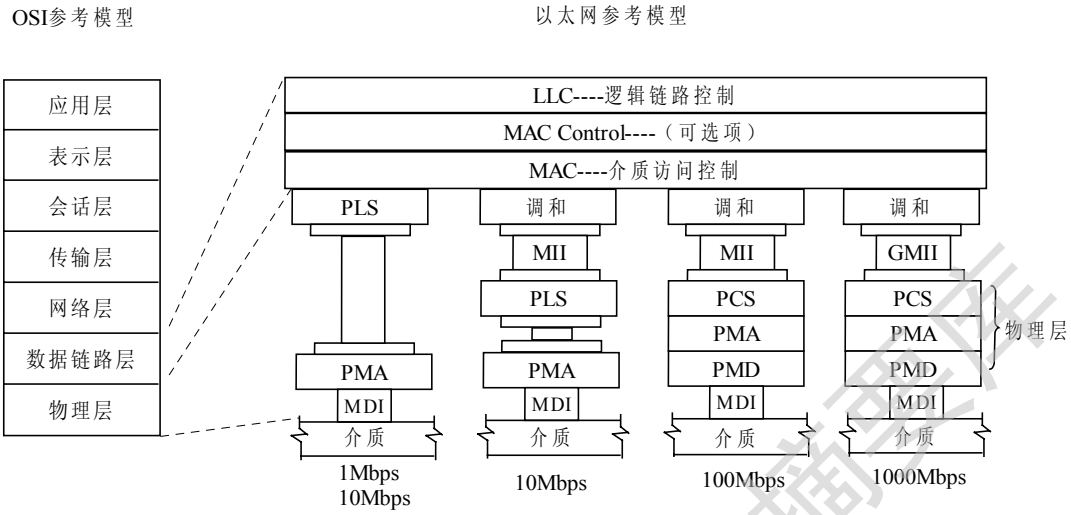
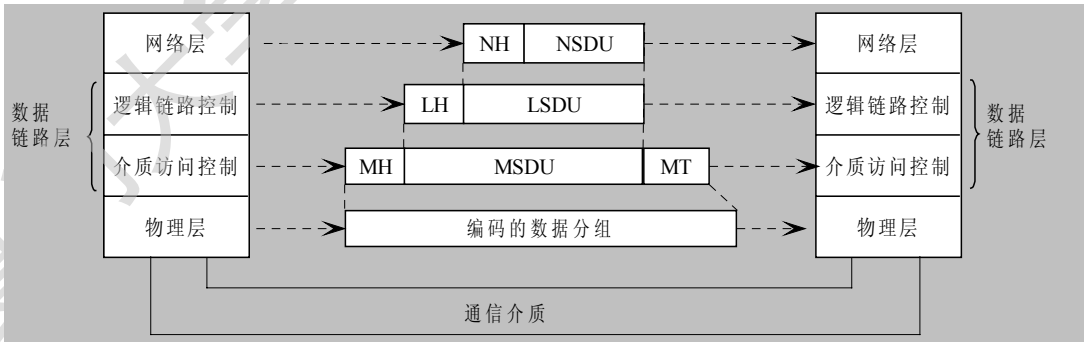


图 1.6 OSI 参考模型与 IEEE802.3 以太网参考模型

在以太网标准中（IEEE802.3）将 OS2 数据链路层的功能分解为一些子层：逻辑链路控制（LLC）和介质访问控制（MAC）和一个可选的 MAC 控制层。同时，OSI 物理层也分解成 3 个子层：物理编码子层（PCS），物理介质附加层（PMA）和物理介质相关子层。

以太网中层与层之间的数据传输路线图如图 1.7：每一层将从上层得到的数据加上自己的报头（也许有报尾），然后将它作为自己的数据发到下一层。直到最后到达物理层，再通过通信介质传到目标站，接收站的每一层依次将自己对应的报头（和报尾）拆卸掉，再指示到上一层。在第一节中所说的帧就是 MAC 层的帧，也就是 MAC 层从上层接收数据片，并给它加上自己的数据头（Preamble,DA,SA,Length/Type）和帧校验序列。然后发给物理编码子层，最后发送到传输介质上。介质另一侧的各层执行相反的工作。



NH = 网络报头　　NSDU = 网络服务数据单元

LH = LLC 报头　　LSDU = 链路服务数据单元

MH = MAC 报头　　MSDU = MAC 服务数据单元

MT = MAC 报尾

图 1.7 分层的数据传输线路

1.3.2 千兆位以太网参考模型介绍

千兆位以太网参考模型如图 1.8。与 100BASE-T 快速以太网相比，由于采用了更快的传输速度，在 MAC 子层定义了两个新特性：载波扩展和帧突发。

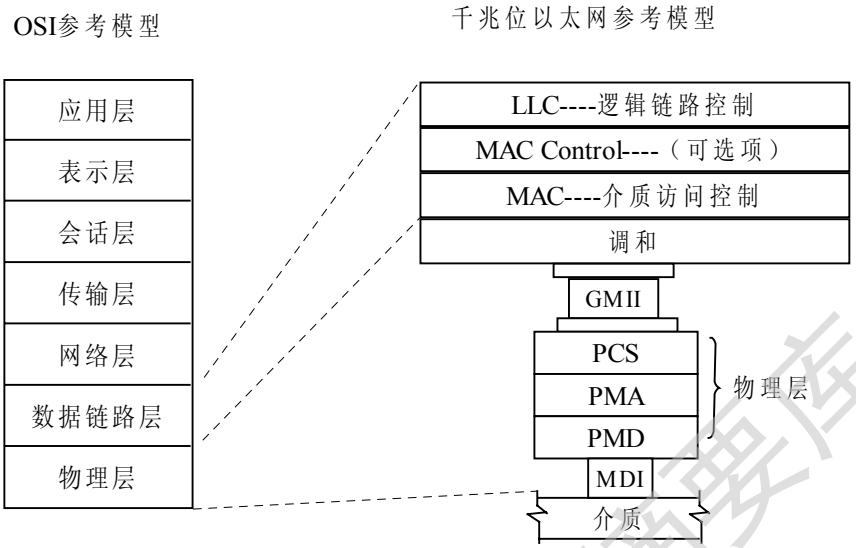


图 1.8 千兆位以太网参考模型

载波扩展：为了保持与低速网络的兼容，仍然定义帧的最小长度位 64 字节。但在千兆位的工作速度下，当使用半双工（CSMA/CD）的方式时，为了得到可接受的网络直径，将 slot Time 定义为 512 字节（4096 位时间）。当发送一个小于 512 字节的短帧时，在 FCS 后附加 0－448 字节的载波扩展。从而保证可以检测到冲突。

帧突发：当某个 DTE 试图发送一个可能需要也可能不需要载波扩展的帧时，同时启动一个突发定时器。如果第一个帧发送完成（即没有发生冲突），而且下述条件成立 DTE 将选择发送另一个网络帧：（i）存在另一个帧要发送，以及（ii）“突发定时器”还未失效。应注意的是突发时间内后序帧不需被扩展。这是因为 solt Time 已经过去，且 DTE 仍控制着介质。一旦突发中止，DTE 将要重新竞争介质控制权。

调和子层（RS）：将 GMII 的八位宽度数据通路及相关控制信号映射到原始 PLS 服务接口定义（MAC/PLS 接口）上。

GMII 接口：提供了千兆位 MAC 和物理层之间的逻辑接口。GMII 和调和子层使得 GMAC 可以连接到不同类型的物理介质上。目前情况下 GMII 接口实际上没有必要，因为所采用的物理介质－光纤和短距离铜介质采用了相同的编码/译码模式。但是在千兆位以太网可以迁移到桌面之前，应当采用 UTP 物理介质之类的低价解决方案。GMII 接口和 MII 接口类似，但作了一些添加/修改。GMII 中的数据通路宽度被指定为 8 位而不是半位元组宽度。每个时钟周期内数据以 1ocset/sec 的速率通过 GMII，因此为了获得 1000Mb/s 的数据速度，发送和接收时钟必须工作在 125MHz.。由于 GMII 接口也可以工作在 10Mb/s 速度，PHY 必须通过站点管理实体来声明字节所支持的速度。以 10Mb/s 速度工作的 GMII 等同于 MII。GMII 接口不是一个外露的接口。GMII 可以用作芯片级的互联（IC 到 IC），这在典型情况下以印刷电路板线路方式，或两个或更多印刷电路板之间的母板到子板连接方式来实现。

GMII 信号：在 GMII 中有一个在以 1000Mb/s 速度工作时使用的独立时钟信号 GTX_CLK。调和子层接口使各种更高速度的 PHY（1000BASE-SX、1000BASE-LX、1000BASE-CX、1000BASE-T 等等）可以连接到千兆位 MAC 引擎之上，并且将来不会有任何升级问题。GMII 由两个独立的数据通路，接收（TXD<7:0>）和发送（TXD<7:0>），各个数据通道的控制信号（RX_ER、RX_DV、TX_ER、TX_EN），网络状态信号（COL、

CRS），各个数据通路的时钟信号（RX_CLK、TX_CLK、GTX_CLK）以及一个两线管理接口（MDC、MDIO）组件。

物理层提供了将数据链路层给出的数据字节转换成可在物理介质上传输的合适信号的手段。同样它也负责将从物理介质上接收到的信号转换成可向数据链路层传递的数据字节。千兆位以太网的物理层技术很大一部分是摘自于 ANSI NCITS T11 光纤通道标准。光纤通道是一种可以千兆位的速度互联工作站，超级计算机，存储设备以及外围设备的技术。在千兆位以太网中采用了最低两层的光纤通道技术，FC-0（接口与介质）和 FC-1（编码/译码）。由于光纤通道技术已经应用了多年，所以 IEEE802.3z 标准委员会为了大大减少千兆位以太网标准的开发时间和风险，而决定采用这种技术。

千兆位以太网的物理层由三个子层构成：

- 物理编码子层（PCS）
- 物理介质接入子层（PMA）
- 物理介质相关子层（PMD）

PCS 位于调和子层（通过 GMII）和物理介质接入（PMA）子层之间。PCS 在 802.3z 标准的第 36 子句中定义。从本质上讲 PCS 完成与 100BASE-X 中相同子层类似的功能，即将经过完善定义的以太网 MAC 功能映射到现存的编码和物理层信号系统的功能上去。100BASE-X 是基于 FD-DI 物理层的。千兆位以太网使用的是经过轻微修改的光纤通道规定。1000BASE-X PCS 采用了规范中的 FC0 和 FC1 部分（光纤通道第 0 和 1 层）。FC-0 定义了包括光纤规范，连接器，以及光电接口在内的物理链路。FC-1 定义了发送编码/解码，错误控制和特殊控制特性。

光纤通道指定了几种数据速率，对于 802.3z 而言，最优的一种是采用 8B/10B 块编码的 1.0625Gb/s 速率。这种编码模式可以使以太网应用达到 850Mb/s（1.0625Gb/s×8/10）的速度。802.3z 成员认为这是不可接受的，并提出将速度增加到 1.25G 波特，以获得 1Gb/s（或 Mb/s）的数据速率。

- 1000BASE-CX——指定工作于两对 150 欧姆平衡式铜电缆上。
- 1000BASE-SX——指定工作使用短波长光的一对光纤上。
- 1000BASE-LX——指定工作使用长波长光的一对光纤上。

图 1.9 中给出了 PCS 的功能模块图。

PMA 子层：提供了 PCS 和 PMD 层之间的串行化服务接口。和 PCS 子层的连接称为 PMA 服务接口。在把 10 位符号传递给 PMD 之前 PMA 子层的发送部分将它们转换成串行位流。而 PMA 子层的接收部分在把接收到的串行位流传递给 PCS 之前把他们转换成 10 位符号。另外 PMA 子层还从接收位流中分离出用于对接收到的数据进行正确的符号对齐（定界）的符号定时时钟。

PMD 子层：其功能是支持在 PMA 子层和介质之间交换串行化的 8B/10B 符号代码位。PMD 子层将这些电信号转换成适合于在某种特定介质上传输的形式。图 1.9 给出了 PMD 子层的功能模块图。PMD 是物理层的最低层，千兆位以太网标准中规定物理层负责从介质上发送和接收信号。IEEE802.3z 标准规定了三种类型的介质：（1）多模光纤上使用 850 纳米激光（1000BASE-SX），（2）单模光纤和多模光纤上使用 1300 纳米激光（1000BASE-LX），（3）短距离铜缆(1000BASE-CX)。另一个独立工作组 IEEE802.3ab，

还定义了一种可以在 UTP 电缆上完成千兆位操作的解决方案，来作为一种桌面互联技术。图 1.10 给出了用于支持千兆位以太网操作的各种不同物理介质。

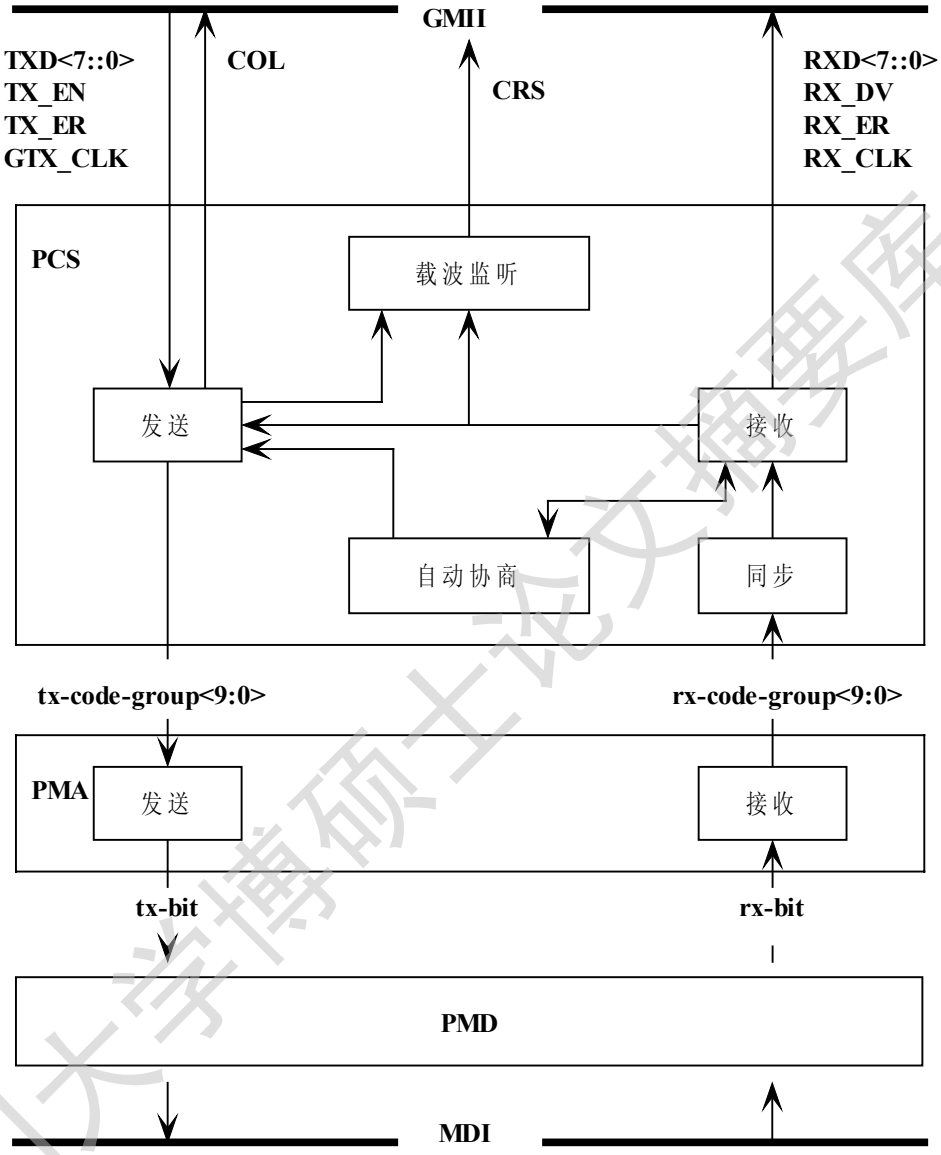


图 1.9 物理层各子层的功能模块图

1.4 半双工千兆位以太网简介

用一个无源星形光纤耦合器（PSC）来取代 802.3z 中定义的中继器，就可以简单的得到一个半双工的千兆位以太网。网络中任一个站点发出的光载波都通过这个 PSC 传输给包括自己在内的所有站点。在一段时间内网络上没有载波，则任意站点都可以开始发送，同时检测冲突。完全按照 CSMA/CD 协议工作。但是由于 PSC 是无源器件，光信号得不到加强，一般只能实现 4－8 个站点的工作组级网络。其功率预算和分配的示意图如图 1.11。

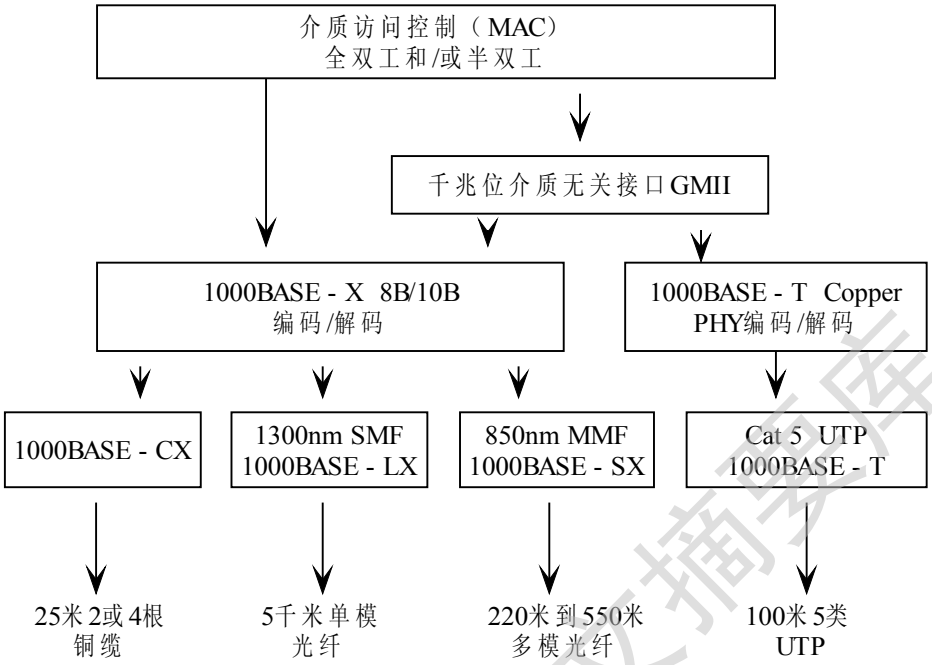


图 1.10 支持千兆位以太网操作的各种不同物理介质

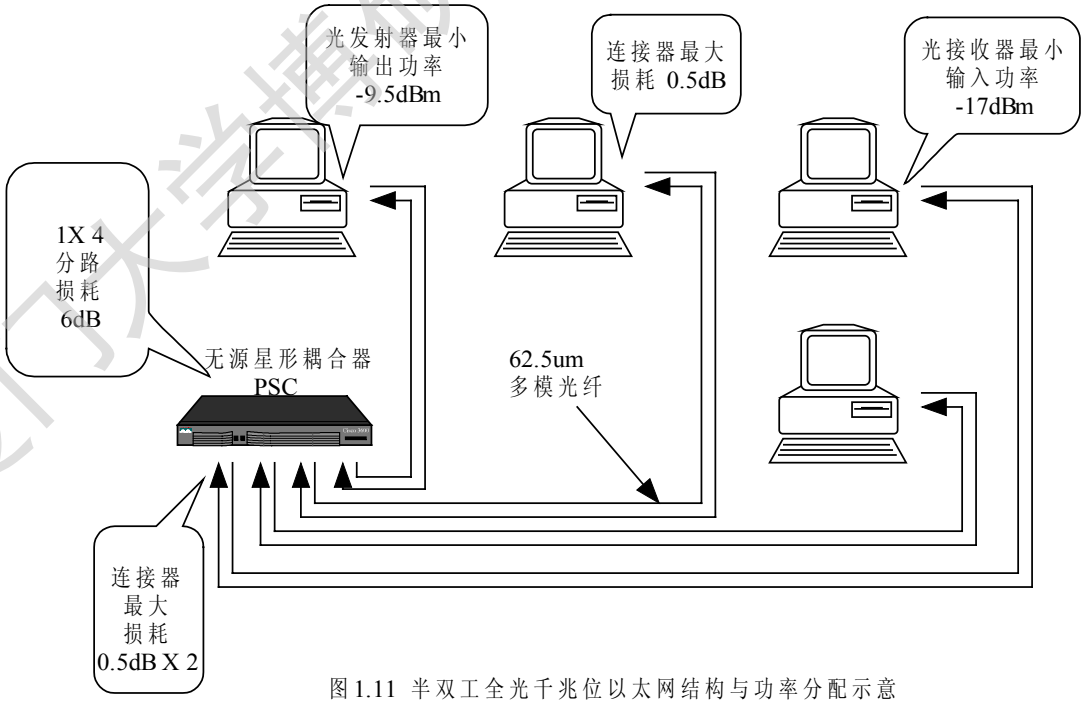


图 1.11 半双工全光千兆位以太网结构与功率分配示意